Deviation Bounds for Wavelet Shrinkage

Dawei Hong

Jean-Camille Birget*

Dept. of Computer Science
Rutgers University at Camden
Camden, NJ 08102, USA
hong@southwest.msus.edu
birget@camden.rutgers.edu

Abstract

We analyse the wavelet shrinkage algorithm of Donoho and Johnstone in order to assess the quality of the reconstruction of a signal obtained from noisy samples. We prove *deviation bounds* for the maximum of the squares of the error, and for the average of the squares of the error, under the assumption that the signal comes from a Hölder class, and the noise samples are independent, of 0 mean, and bounded. Our main technique is Talgrand's isoperimetric theorem. Our bounds refine the known *expectations* for the average of the squares of the error.

*Second author's research supported in part by NSF grant DMS-9970471

1 Introduction

We address the classical problem of the reconstruction of signal samples from noisy samples. We consider an original *signal* of bounded duration $f\colon\ t \in [0,1] \to f(t) \in \mathbb{R}$. We also have *additive noise* $e\colon [0,1] \to \mathbb{R}$. Thus, the observed *noisy signal* at time t is $y(t) = f(t) + e(t)$.

We sample the noisy signal at n uniformly spaced instants and we denote the sample values by $y_i = f_i + e_i = f(\frac{i}{n}) + e(\frac{i}{n})$ (for $1 \le i \le n$). Our goal is to recover a good approximation of the original signal samples $(f_1, \ldots, f_n)$ from the noisy signal samples $(y_1, \ldots, y_n)$. For this to be possible we need some assumptions that distinguish the signal from the noise:

- The original signal f has a certain degree of "smoothness", i.e., f belongs to a Hölder class $\Lambda^\alpha(M)$ for some $\alpha > 0$ and $M > 0$.
- The noise is "random", i.e., $(e_1, \ldots, e_n)$ consists of n independent Borel random variables.

The *Hölder classes* are defined as follows:

For $0 < \alpha \le 1$, $\Lambda^\alpha(M) = \{h \in \mathbb{R}^{[0,1]} : (\forall x_1, x_2 \in [0,1]),\ |h(x_1) - h(x_2)| \le M|x_1 - x_2|^\alpha\}$.

For $1 < \alpha$, $\Lambda^\alpha(M) = \{h \in \mathbb{R}^{[0,1]} : (\forall x \in [0,1])\ |h'(x)| \le M,\ h^{\lfloor\alpha\rfloor}$ exists, and $(\forall x_1, x_2 \in [0,1])\ |h^{\lfloor\alpha\rfloor}(x_1) - h^{\lfloor\alpha\rfloor}(x_2)| \le M|x_1 - x_2|^{\alpha - \lfloor\alpha\rfloor}\}$.

Let $(\tilde{y}_1, \ldots, \tilde{y}_n)$ be an approximation of $(f_1, \ldots, f_n)$, obtained from $(y_1, \ldots, y_n)$. Most commonly, the closeness of this approximation is measured by $\frac{1}{n}\sum_{i=1}^n(\tilde{y}_i - f_i)^2$ or by the expectation $\mathbf{E}[\frac{1}{n}\sum_{i=1}^n(\tilde{y}_i - f_i)^2]$ (which makes sense since the e_i, and hence the $\tilde{y}_i$, are random variables).

The **wavelet shrinkage algorithm** of Donoho and Johnstone [6], [7] is a very efficient tool for finding good estimates $\tilde{y}$. In outline, the algorithm works as follows:

(Step 0) Choose a wavelet system with N vanishing moments ($N \ge \alpha$); choose a level of coarseness $J_0 \ge 0$ (J_0 will depend on α), and consider the multi-resolution chain of Hilbert spaces $V_{J_0} \subset V_{J_0+1} \subset \ldots \subset V_j \subset \ldots$.

(Step 1) Apply the *Discrete Wavelet Transform* (DWT) to the noisy signal samples $(y_1, \ldots, y_n)$, where $n \ge 2^{J_0}$. This yields the "empirical wavelet coefficients" $(\xi_1, \ldots, \xi_n)$.

(Step 2) Fix a "threshold" λ_n (> 0) and apply either "hard" or "soft thresholding" to $(\xi_1, \ldots, \xi_n)$.

Hard thresholding consists of replacing each ξ_i by 0 when $|\xi_i| \le \lambda_n$, and keeping ξ_i unchanged when $|\xi_i| > \lambda_n$.

Soft thresholding consists of transforming each ξ_i as follows: ξ_i is replaced by 0 if $|\xi_i| \le \lambda_n$; if $\xi_i > \lambda_n$, ξ_i is replaced by $\xi_i - \lambda_n$; if $\xi_i < -\lambda_n$, ξ_i is replaced by $\xi_i + \lambda_n$.

(Step 3) Apply the inverse DWT to the result of (2). This yields the estimate $(\tilde{y}_1, \ldots, \tilde{y}_n)$.

To what extent does wavelet shrinkage depend on the smoothness conditions of the signal f and on the randomness conditions of the noise samples e_i, and how do the estimators $\tilde{y}_i$ approximate the original signal f? In [6], [7] it was assumed that the e_i are iid Gaussian variables with distribution $N(0, \sigma^2)$, and the threshold was chosen to be $\lambda_n = \sigma\sqrt{2\frac{\log n}{n}}$. Assuming that $f \in \Lambda^\alpha(M)$ (the Hölder class) with $\alpha > 0$, it is proved in [6], [7] that $\mathbf{E}[\frac{1}{n}\sum_{i=1}^n(\tilde{y}_i - f_i)^2] <$

$C \cdot (\frac{1}{n}\log n)^{\frac{2\alpha}{1+2\alpha}}$, where C depends only on M and on the wavelet system used. It was observed in [6], [7] (the proofs are due to Lepskii [9] and to Brown and Low [3]) that this upper bound is optimal over all possible algorithms, if the parameters α and M are *not* known. For the optimality of the wavelet shrinkage algorithm it is important that the threshold be of the form $c \cdot \sqrt{\frac{\log n}{n}}$ (where c does not depend on n).

Since the publication of [6], [7] there has been further progress on wavelet shrinkage (chapter 6 of [13] is an excellent reference up to 1999). Most recently, Averkamp and Houdré [1], [2] expanded the scope of wavelet shrinkage by allowing the noise samples e_i to have different distributions F_i, chosen from a wide class of distributions. They show in [1] (page 32) that the error expectation of the wavelet shrinkage algorithm for bounded noise is roughly the same as for Gaussian noise, if the parameters α and M of the Hölder class of the signal are not known. They also discuss various choices of thresholds.

All the results on wavelet shrinkage in the literature so far evaluate the quality of the approximation by bounding the expectation $\mathbf{E}[\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - f_i)^2]$, to the best of our knowledge. In this paper we study deviation bounds (rather than just the expectation) of $\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - f_i)^2$ and of $\max\{(\tilde{y}_i - f_i)^2 : 1 \le i \le n\}$.

Assumptions: We assume that the signal f belongs to a Hölder class $\Lambda^\alpha(M)$, and that the noise samples e_i are *independent* random variables (with possibly different distributions). The only restrictions on the distributions are that they are Borel measurable, have *compact support* (contained in an interval $[-\frac{b}{2}, \frac{b}{2}]$), and zero mean. The assumption that the distributions of the noise have bounded support is of course equivalent to assuming that the noise e_i has bounded values ($|e_i| \le \frac{b}{2}$).

The main results of this paper are the following deviation bounds.

Theorem. *For the wavelet shrinkage algorithm with threshold*

$$\lambda_{n,\delta} = C_\varphi\, b\,(1 + 2\sqrt{(1+\delta)\ln 2}\,)\sqrt{\tfrac{\log n}{n}}$$

(where C_φ depends only on the wavelet system) we have the following deviation bounds: There are $c_1, c_2 > 0$, depending only on b, M, and α, such that for all $n \ge n_0$ and all $\delta > 0$,

$$\mathbf{P}\left(max\{(\tilde{y}_i - f_i)^2 : 1 \le i \le n\} \;\le\; (c_1 + c_2\delta)\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{1+2\alpha}} \right) \;\ge\; 1 - \frac{9}{n^{1+\delta}}.$$

As a consequence,

$$\mathbf{P}\left(\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - f_i)^2 \;\le\; (c_1 + c_2\delta)\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{1+2\alpha}} \right) \;\ge\; 1 - \frac{9}{n^{1+\delta}}.$$

The minimum number of samples, n_0, is 2^9 when $0 < \alpha \le 1$; when $\alpha > 1$, $n_0 = \; (4\alpha+2)^{2\alpha+2} \cdot (\log_2(4\alpha+2))^2$.

One notices that n_0 grows very rapidly with α, when $\alpha > 1$. For $\alpha = 2$, we have $n_0 = 1.1 * 10^7$; for $\alpha = 3$, $n_0 = 3.7 * 10^{10}$, which is impractical. So for large α our theorem is interesting only from an asymptotic point of view. On the other hand, in practice usually $\alpha \le 1$.

2 Preliminaries

2.1 Wavelets

We will usually follow the notation of [5] regarding wavelets, the only exception being that we reverse the multi-resolution indices. Moreover, we only consider real-valued functions with domain $[0,1]$. So we have a sequence of real Hilbert spaces $V_{J_0} \subset V_{J_0+1} \subset \ldots \subset V_j \subset \ldots$, such that the closure of $\bigcup_j V_j$ is $\mathrm{L}^2[0,1]$. We let $V_{j+1} = V_j \oplus W_j$ (orthogonal complement). Since we are in the case of compactly supported functions each V_j is a finite-dimensional real vector space (of dimension 2^j), with orthonormal basis $\{\varphi_{j,k} : 0 \le k \le 2^j - 1\}$, derived from a scaling function φ. Let ψ be the wavelet function corresponding to φ, and let $\{\psi_{j,k} : 0 \le k \le 2^j - 1\}$ be the corresponding orthonormal basis of W_j.

For any function $g \in \mathrm{L}^2[0,1]$ we define the piece-wise constant function $\overline{g}$: $[0,1] \to \mathbb{R}$ as follows: $\overline{g}(x) = g(\frac{k}{n})$ $(= g_k)$ if $\frac{k-1}{n} < x \le \frac{k}{n}$ for some $k = 1, \ldots, n$; $\overline{g}(x) = 0$ if $x \notin]0,1]$. The discrete wavelet transform of a vector $(g_1, \ldots, g_n)$ can be obtained by taking the wavelet coefficients of the piecewise constant function $\overline{g}$. These *wavelet coefficients* are:

$c_{j,k}^{(g)} = \langle \overline{g}, \varphi_{j,k} \rangle = \int_0^1 \overline{g}(x)\, \varphi_{j,k}(x)\, dx$, and

$d_{j,k}^{(g)} = \langle \overline{g}, \psi_{j,k} \rangle = \int_0^1 \overline{g}(x)\, \psi_{j,k}(x)\, dx$.

Then for any integer $J \ge J_0$:

$$\overline{g}(x) = \sum_{k=0}^{2^J-1} c_{J,k}^{(g)} \varphi_{J,k}(x) + \sum_{j=J}^{+\infty} \sum_{k=0}^{2^j-1} d_{j,k}^{(g)} \psi_{j,k}(x) \qquad \textbf{a.e.}$$

In this paper we will use two wavelet systems: The Haar wavelets (because of their simplicity, especially for programming purposes), and the interval wavelets with predefined vanishing moments, based on Daubechies wavelets (Cohen, Daubechies, Jawerth, Vial [4]).

For the *Haar wavelets*, the scaling function is $\varphi(x) = 1$ when $0 < x \le 1$, and $\varphi(x) = 0$ otherwise. Hence, $\varphi_{j,k}(x) = 2^{j/2}$ when $k2^{-j} < x \le (k+1)2^{-j}$, and $\varphi_{j,k}(x) = 0$ otherwise. The Haar wavelet function is $\psi(x) = 1$ if $0 < x \le \frac{1}{2}$, $\psi(x) = -1$ if $\frac{1}{2} < x \le 1$, and $\psi(x) = 0$ otherwise. Hence, $\psi_{j,k}(x) = 2^{j/2}$ if $k2^{-j} < x \le (k + \frac{1}{2})2^{-j}$, $\psi_{j,k}(x) = -2^{j/2}$ if $(k+\frac{1}{2})2^{-j} < x \le (k+1)2^{-j}$, and $\psi_{j,k}(x) = 0$ otherwise.

For the *interval wavelet system* of [4], with N *vanishing moments*, the scaling function φ and the wavelet function ψ are complicated. But all we need to know about them is the following:

- A multiresolution of $L^2[0,1]$ is obtained, with an orthonormal basis for V_j when $j > J_0$:

  $\{\varphi_{j,k} : 1 \le k < 2^j - 2N\} \cup \{\varphi_{j,i}^{\text{left}}, \varphi_{j,i}^{\text{right}} : 0 \le i < N\}$.

Each $\varphi_{j,k}$ has support $[k2^{-j}, (2N-1+k)2^{-j}]$, each $\varphi_{j,i}^{\text{left}}$ has support $[0, i2^{-j}]$, and each $\varphi_{j,i}^{\text{right}}$ has support $[1 - i2^{-j}, 1]$.

The decomposition level J_0 is chosen so that $J_0 \ge 1 + \log_2(2N-1)$. For signals in the Hölder class $\Lambda^\alpha(M)$ we require the number of vanishing moments to be $N \ge \alpha$.

- We also have an orthonormal basis for W_j,

  $\{\psi_{j,k} : 1 \le k < 2^j - 2N\} \cup \{\psi_{j,i}^{\text{left}}, \psi_{j,i}^{\text{right}} : 0 \le i < N\}$

with the same supports as the corresponding φ functions.

• φ and ψ are bounded on $[0,1]$ by a constant $C > 0$, independent of x and N: $\forall x \in [0,1]$, $|\varphi(x)|, |\psi(x)| \leq C$.

For $0 \leq k < 2^j - 2N$ ("inside the the interval"), $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$.
At the ends of the interval $[0,1]$ we have for $0 \leq i < N$, (see [4])
$\varphi^{\text{left}}_{j,i}(x) \;=\; \sum_{h=1}^{2N-1}(-h)^i\varphi(2^j x + h)$.
A similar formula holds on the right end of the interval $[0,1]$.

Assuming that n is a power of 2, $n = 2^J$, we have for the function $\overline{y}$, relative to any wavelet system: $\overline{y}(x) = \sum_{k=0}^{2^J-1}\langle \overline{y}, \varphi_{J,k}\rangle \varphi_{J,k}(x)$. Thus for any J_1 with $0 \leq J_1 < J$, the DWT transforms $(y_1, \ldots, y_n)$ to $\sqrt{n}\,(c^{(\overline{y})}_{J_1,0}, \ldots,\, c^{(\overline{y})}_{J_1,2^J-1},\, d^{(\overline{y})}_{J_1,0}, \ldots, d^{(\overline{y})}_{J_1,2^J-1},\, \ldots, \ldots,\, d^{(\overline{y})}_{J-1,0}, \ldots, d^{(\overline{y})}_{J-1,2^J-1})$. The DWT is an orthogonal transformation (represented by an orthogonal matrix W).

We will always assume that n is a power of 2: $n = 2^J$. Throughout this paper, log will refer to $\log_2$, and ln will denote the natural logarithm.

Let us now return to the analysis of a noisy signal $y(t) = f(t) + e(t)$.

Lemma 2.1 *With respect to the Haar wavelets, the wavelet coefficients of the function e have the following properties:*

(H1) For all $j \in [0, 2^J]$ and all $k \in [0, 2^{j-1}-1]$:

$$c^{(e)}_{j,k} = 2^{-J+j/2}\sum_{i=0}^{2^{J-j}-1} e_{i+1+k2^{J-j}}$$

(H2) For all j and k as in (H1):

$$d^{(e)}_{j,k} = 2^{-J+j/2}\sum_{i=0}^{2^{J-j-1}-1}(e_{i+1+k2^{J-j}} - e_{i+1+(k+\frac{1}{2})2^{J-j}})$$

For any function $f : [0,1] \to \mathbb{R}$ belonging to $\Lambda^{(\alpha)}(M)$ with $0 < \alpha \leq 1$ we have:

(H3) For all $j \in [0, 2^J]$ and all $k \in [0, 2^{j-1}-1]$:

$$|d^{(f)}_{j,k}| \;<\; M\,2^{-j(\frac{1}{2}+\alpha)}.$$

The proof of this lemma is just a calculation and is given in the Appendix.

Lemma 2.2 *With respect to the interval wavelet system [4], the wavelet coefficients of the function e have the following properties:*

(D1) *For all $j \in [0, 2^J]$ and all $k \in [0, 2^{j-1} - 1]$:*

$$c_{j,k}^{(e)} = 2^{-J+j/2} \sum_{i=0}^{2^{J-j}-1} \alpha_{i,j,k} e_{i+1+k2^{J-j}}$$

for some numbers $\alpha_{i,j,k}$ that do not depend on the noise function e. Moreover, $|\alpha_{i,j,k}| < C_\varphi$ for some constant $C_\varphi \geq 1$ depending only on the wavelet system.

(D2) *For all j and k as in (D1):*

$$d_{j,k}^{(e)} = 2^{-J+j/2} \sum_{i=0}^{2^{J-j}-1} \beta_{i,j,k} e_{i+1+k2^{J-j}}$$

for some numbers $\beta_{i,j,k}$ that do not depend on the noise function e. Moreover, $|\beta_{i,j,k}| < C_\varphi$ where $C_\varphi \geq 1$ depends only on the wavelet system.

Suppose $f : [0,1] \to \mathbb{R}$ belongs to $\Lambda^{(\alpha)}(M)$ with $1 < \alpha$, and suppose the number of vanishing moments N of the wavelet system satisfies $N \geq \alpha$. Then we have:

(D3) *For all $j \in [0, 2^J]$ and all $k \in [0, 2^{j-1} - 1]$:*

$$|d_{j,k}^{(f)}| \; < \; C_\varphi \, M \, 2^{-j(\frac{1}{2}+\alpha)}$$

where $C_\varphi \geq 1$ depends only on the wavelet system.

The proof of Lemma 2.2 is just a calculation and is given in the Appendix.

2.2 Talagrand's isoperimetric theorems

Talagrand's isoperimetric theorems, published in 1995 [12], have had a profound impact on the probabilistic analysis of combinatorial optimization methods; Talagrand's theorems often apply quite directly, giving shorter proofs, often with dramatically better results than previously used methods (see [11], chapter 6). We will use the following result of [12].

Let (Ω, Σ, μ_i) $(i = 1, \ldots, n)$ be Borel probability spaces, and let Ω^n be the product space with product measure $P = \mu_1 \times \ldots \times \mu_n$. For $A \subseteq \Omega^n$ and $\omega = (\omega_1, \ldots, \omega_n) \in \Omega^n$, *Talagrand's 'convex' distance* is defined by

$$d_T(\omega, A) = \sup \left\{ \inf \left\{ \sum_{i=1}^{n} \beta_i \cdot I(\omega_i \neq a_i) : (a_1, \ldots, a_n) \in A \right\} : (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n, \sum_{i=1}^{n} \beta_i^2 = 1 \right\}.$$

Notation: $I(\omega_i \neq a_i) = 1$ if $\omega_i \neq a_i$, and $I(\omega_i \neq a_i) = 0$ otherwise.

Theorem 2.3 *(Talagrand, Theorem 4.1.1 in [12]): For any $A \subseteq \Omega^n$ with $P(A) > 0$:*

$$\int_{\Omega^n} \exp(\frac{1}{4} d_T(\omega, A)^2) dP(\omega) \ \leq \ \frac{1}{P(A)}.$$

As a corollary, for all $t > 0$,

$$P(d_T(\omega, A) \geq t) \ \leq \ \frac{1}{P(A)} \cdot \exp(-\frac{t^2}{4}).$$

3 Deviation bound for $\frac{1}{n}\sum_{i=1}^{n}(f_i - \tilde{y}_i)^2$

Recall that the input for wavelet shrinkage is $(y_1, \ldots, y_n)$, where $y_i = f_i + e_i \ \ (i = 1, \ldots, n)$, the f_i are samples from the original signal f, and the e_i are additive noise. The e_i are independent Borel random variables. We assume that the noise is bounded (with $|e_i| \leq \frac{b}{2}$), so each random variable e_i is a Borel measurable function e_i: $\omega_i \in \Omega \mapsto e_i(\omega_i) \in [-\frac{b}{2}, \frac{b}{2}]$. Accordingly, we view $(e_1, \ldots, e_n)$ as a function $\omega = (\omega_1, \ldots, \omega_n) \in \Omega^n \ \mapsto \ e(\omega) = (e_1(\omega_1), \ldots, e_n(\omega_n)) \in [-\frac{b}{2}, \frac{b}{2}]^n$. (Borel measurability is assumed in order to apply Talagrand's theorem.) To simplify the notation we often write $e_i(\omega)$ for $e_i(\omega_i)$.

We shall first define a subset A of Ω^n and then show that

- $P(A) > \frac{1}{9}$ if n is large enough, and
- wavelet shrinkage satisfies our deviation bounds when the noise samples are in A.

Then for any $\delta > 0$ we define a subset $B_\delta \subseteq \Omega^n$ such that

- for any $\omega \in \Omega^n$, if Talagrand's distance satisfies $d_T(\omega, A) \leq 2\sqrt{(1+\delta)\ln n}$ then $\omega \in B_\delta$;
- wavelet shrinkage satisfies our deviation bounds when the noise samples are in B_δ.

Finally, by applying Talagrand's theorem we obtain our results.

3.1 The subset A

Recall that we assume $n = 2^J$. For any $\omega \in \Omega^n$ we decompose the noise sample sequence $e(\omega)$ into blocks of length J, as follows:

$e(\omega) = (\ldots, \ \ldots, \ e_{kJ+1}(\omega), \ldots, e_{(k+1)J}(\omega), \ \ldots, \ \ldots)$

where $k = 0, \ \ldots, \ \frac{1}{J}2^J - 1$. Here, for simplicity we regard $\frac{1}{J}2^J = 2^{J - \log J}$ as an integer (i.e., we assume that J is a power of 2).

For the Haar wavelets we define the subset $A \subset \Omega^n$ as follows:

$$A \ = \ \{\omega \in \Omega^n : (\forall \ell \in [-1, J - \log J])(\forall k \in [0, 2^{J - \log J - \ell} - 1]),$$

$$\left|\sum_{i=0}^{J2^{\ell-1}-1} e_{k2^\ell J+i+1}(\omega)\right| \leq bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \}.$$

For the interval wavelet system we define

$$\begin{aligned} A \ &= \ \{\omega \in \Omega^n : (\forall \ell \in [-1, J-\log J])(\forall k \in [0, 2^{J-\log J-\ell}-1]), \\ &\left|\sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega)\cdot \alpha_{i,J-\log J-\ell,k}\right| \leq bJ2^{\ell/2}\sqrt{2^{-1}\ln 2} \\ \text{and} \quad &\left|\sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega)\cdot \beta_{i,J-\log J-\ell,k}\right| \leq bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \}. \end{aligned}$$

We need a classical result from probability theory.

Theorem 3.1 *(Hoeffding's inequality) Let $X_1, \ldots, X_m$ be independent random variables with $b_1 \leq X_i \leq b_2 \quad (i = 1, \ldots, m)$. Then for all $t > 0$,*

$$P\left(\left|\sum_{i=1}^{m}(X_i - E[X_i])\right| \leq t\right) \geq 1 - \exp\left(-\frac{2t^2}{m(b_2-b_1)^2}\right).$$

Lemma 3.2 *For all $n > 1$, $P(A) \geq 1 - \frac{4}{\log n} + \frac{1}{n}$ for the Haar wavelets, and $P(A) \geq 1 - \frac{8}{\log n} + \frac{2}{n}$ for the interval wavelet system.*

In either case, if $n \geq 256$ then $P(A) \geq \frac{1}{128}$. If $n \geq 2^9$ then $P(A) > \frac{1}{9}$. Moreover, $P(A)$ tends to 1 when $n \to \infty$.

Proof: We first give the proof for the Haar wavelets. For any $\ell \in [-1, J-\log J]$ and $k \in [0, 2^{J-\log J-\ell}-1]$ the noise samples $e_{k2^\ell J+1}, \ldots, e_{(k+1)2^\ell J}$ are independent random variables, each with values in $[-\frac{b}{2}, \frac{b}{2}]$. So Hoeffding's inequality applies, and since $E[e_i] = 0$ for all i, we obtain for all $t > 0$,

$$P\left(\left|\sum_{i=0}^{2^{\ell-1}J-1} e_{k2^\ell J+i+1}\right| \leq t\right) \ \geq \ 1 - \exp\left(-\frac{2t^2}{2^\ell J b^2}\right).$$

Letting $t = b2^{\ell/2}J\sqrt{2^{-1}\ln 2}$ we obtain

$$P\left(\left|\sum_{i=0}^{2^{\ell-1}J-1} e_{k2^\ell J+i+1}\right| \leq b2^{\ell/2}J\sqrt{2^{-1}\ln 2}\right) \geq 1 - \frac{1}{n}. \tag{1}$$

For $\ell \in [-1, J - \log J]$ and $k \in [0, 2^{J-\log J-\ell} - 1]$, let

$$A_{\ell,k} = \left\{ \omega \in \Omega^n \;:\; \left| \sum_{i=0}^{2^{\ell-1}J-1} e_{k2^\ell J+i+1}(\omega) \right| \le b2^{\ell/2} J \sqrt{2^{-1}\ln 2} \right\}$$

and let $A_\ell = \bigcap_{k=0}^{2^{J-\log J-\ell}-1} A_{\ell,k}$.
Then by (1), $P(A_{\ell,k}) \ge 1 - \frac{1}{n}$.
For the complements of these sets we have $\bar{A}_\ell = \bigcup_{k=0}^{2^{J-\log J-\ell}-1} \bar{A}_{\ell,k}$
hence $P(\bar{A}_\ell) \le \sum_{k=0}^{2^{J-\log J-\ell}-1} \frac{1}{n}$.
Since $n = 2^J$ we obtain $P(\bar{A}_\ell) \le \frac{2^{-\ell}}{\log n}$.
Since $A = \bigcap_{\ell=-1}^{J-\log J} A_\ell$ we have

$P(A) \ge \; 1 - \sum_{\ell=-1}^{J-\log J} P(\bar{A}_\ell) \;\ge\; 1 - \sum_{\ell=-1}^{J-\log J} \frac{2^{-\ell}}{\log n}$.

Hence, $P(A) \ge 1 - \frac{4}{\log n} + \frac{1}{n}$. This proves the Lemma for the Haar case.

For the interval wavelet system we let

$$\begin{aligned} A^\alpha \;=\; & \{\omega \in \Omega^n \;:\; (\forall \ell \in [-1, J - \log J])(\forall k \in [0, 2^{J-\log J-\ell} - 1]), \\ & \left| \sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega) \cdot \alpha_{i,J-\log J-\ell,k} \right| \le bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \}, \end{aligned}$$

and

$$\begin{aligned} A^\beta \;=\; & \{\omega \in \Omega^n \;:\; (\forall \ell \in [-1, J - \log J])(\forall k \in [0, 2^{J-\log J-\ell} - 1]), \\ & \left| \sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega) \cdot \beta_{i,J-\log J-\ell,k} \right| \le bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \}. \end{aligned}$$

Then $A = A^\alpha \cap A^\beta$.
We also let

$$A^\alpha_{\ell,k} = \{\omega \in \Omega^n \;:\; \left| \sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega) \cdot \alpha_{i,J-\log J-\ell,k} \right| \le bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \},$$

and

$$A^\beta_{\ell,k} = \{\omega \in \Omega^n \;:\; \left| \sum_{i=0}^{J2^\ell-1} e_{k2^\ell J+i+1}(\omega) \cdot \beta_{i,J-\log J-\ell,k} \right| \le bJ2^{\ell/2}\sqrt{2^{-1}\ln 2}\ \}.$$

Moreover, we let $A^\alpha_\ell = \bigcap_k A^\alpha_{\ell,k}$ and $A^\beta_\ell = \bigcap_k A^\beta_{\ell,k}$. Then $A_\ell = A^\alpha_\ell \cap A^\beta_\ell$, hence $\bar{A}_\ell = \bar{A}^\alpha_\ell \cup \bar{A}^\beta_\ell$.

By the same proof as for Haar wavelets above: $P(\bar{A}_\ell^\alpha)$ and $P(\bar{A}_\ell^\beta) \le \frac{2^{-\ell}}{\log n}$.

Hence, $P(\bar{A}_\ell) \le \frac{2^{-\ell+1}}{\log n}$.

Since $A = \bigcap_{\ell=-1}^{J-\log J} A_\ell$ we obtain by a similar calculation as in the Haar case:

$P(A) \ge 1 - \frac{8}{\log n} + \frac{2}{n}$. $\square$

Lemma 3.3 *For all $\omega \in A$, all $j \in \,]J_0, J[$, and all $k \in [0, 2^j - 1]$, we have (for some constant $C_\varphi \ge 1$, depending only on the wavelet system):*

$$|d_{j,k}^{(e(\omega))}| \le b\, C_\varphi \sqrt{\frac{\log n}{n}}$$

and for all $k \in [0, 2^{J_0} - 1]$,

$$|c_{J_0,k}^{(e(\omega))}| \le b\, C_\varphi \sqrt{\frac{\log n}{n}}$$

Proof: We consider two cases for j.

Case 1: $J_0 \le j \le J - \log J + 1$.

We write j as $J - \log J - \ell$, where $-1 \le \ell \le J - \log J - J_0$. Let us first consider Haar wavelets. By (H2) (in Lemma 2.1) we have

$$d_{j,k}^{(e(\omega))} = 2^{-J+j/2} \left(\sum_{i=0}^{2^{\ell-1}J-1} e_{k2^\ell J+i+1}(\omega) - \sum_{i=0}^{2^{\ell-1}J-1} e_{(k+1/2)2^\ell J+i+1}(\omega) \right).$$

Since $\omega \in A$ we can apply the defining property of A to

$\left|\sum_{i=0}^{J2^{\ell-1}-1} e_{i+1+k2^\ell J}\right| = \left|\sum_{i=0}^{J2^{\ell-1}-1} e_{i+1+2k2^{\ell-1}J}\right|$.

Since $2k$ is in the correct range $[0, 2^{j+1} - 2] = [0, \frac{1}{J}2^{J-(\ell-1)} - 2]$, we have

$\left|\sum_{i=0}^{J2^{\ell-1}-1} e_{i+1+k2^\ell J}\right| \le bJ2^{(\ell-1)/2}\sqrt{2^{-1}\ln 2}$.

Similarly,

$\left|\sum_{i=0}^{J2^{\ell-1}-1} e_{i+1+(k+\frac{1}{2})2^\ell J}\right| = \left|\sum_{i=0}^{J2^{\ell-1}-1} e_{i+1+(2k+1)2^{\ell-1}J}\right| \le bJ2^{(\ell-1)/2}\sqrt{2^{-1}\ln 2}$;

we used the defining property of A, since the range of $2k+1$ is

$[0, 2^{j+1} - 2 + 1] = [0, \frac{1}{J}2^{J-(\ell-1)} - 1]$.

By combining these two bounds we obtain

$|d_{j,k}^{(e(\omega))}| \;\le\; 2^{-J+j/2} \cdot 2 \cdot bJ2^{(\ell-1)/2}\sqrt{2^{-1}\ln 2} \;<\; b\sqrt{\ln 2}\sqrt{\frac{\log n}{n}} \;\le\; b\sqrt{\frac{\log n}{n}}$.

Let us now consider case 1 for the interval wavelet system. By (D2) in Lemma 2.2,

$d_{j,k}^{(e(\omega))} = 2^{-J+j/2} \cdot \sum_{i=0}^{2^\ell J-1} e_{k2^\ell J+i+1}(\omega) \cdot \beta_{i,j,k}$.

Since $\omega \in A$,

$$|d_{j,k}^{(e(\omega))}| \;\le\; 2^{-J+j/2} \cdot bJ2^{(\ell-1)/2}\sqrt{2^{-1}\ln 2} \;=\; b2^{(-J+\log J)/2}\sqrt{2^{-1}\ln 2} \;=\; b\sqrt{\tfrac{\log n}{n}}\sqrt{2^{-1}\ln 2}$$
$$\le\; b\sqrt{\tfrac{\log n}{n}}.$$

Case 2: $J - \log J + 2 \le j < J$.
For the Haar wavelets we use the boundedness of the noise, $|e_i - e_j| \le b$. Hence, by (H2),

$$|d_{j,k}^{(e(\omega))}| \le 2^{-J+j/2} b(J2^{\ell-1} - 1) \;\le\; b\sqrt{\tfrac{\log n}{n}}.$$

For the interval wavelet system, (D2) yields

$$|d_{j,k}^{(e(\omega))}| \le\; 2^{-J+j/2} \sum_{i=0}^{2^{J-j}-1} |e_{k2^{\ell}J+i+1}(\omega)| \cdot |\beta_{i,j,k}| = 2^{-J+j/2}\, 2^{J-j}\, \tfrac{b}{2}\, C_\varphi$$
$$\le\; \tfrac{b}{2}\, C_\varphi\, 2^{-j/2} \;\le\; b\, C_\varphi \sqrt{\tfrac{\log n}{n}}$$

by using $j \ge J - \log J + 2$ for the last inequality.

By an argument similar to the above we obtain the bound for $|c_{J_0,k}^{(e(\omega))}|$. $\square$

To implement wavelet shrinkage we need two parameters: A decomposition level J_0 and a threshold $\lambda_{n,\delta}$. We define

$$J_1 = \lceil \frac{1}{1+2\alpha}(J - \log J) \rceil$$

and we choose J_0 so that $J_0 \le J_1$.

For the Haar wavelets (when $0 < \alpha \le 1$) we can simply pick $J_0 = 0$, but for the interval wavelet system (when $1 < \alpha$ and we have $N = \lceil \alpha \rceil$ vanishing moments), we also require (see [4]) that $J_0 \ge 1 + \log(2N - 1)$. When $\alpha > 1$ we choose

$$J_0 = 1 + \lceil \log(2\lceil \alpha \rceil - 1) \rceil$$

Thus, for J_0 to exist (when $\alpha > 1$) we need $n = 2^J$ to be such that $1 + \log(2\lceil\alpha\rceil - 1) \le J_1$. A sufficient condition for this is that $J - \log J \ge (1 + \log(2\alpha + 1))\,(1 + 2\alpha)$,
or equivalently, $\frac{n}{\log n} \ge (4\alpha + 2)^{2\alpha+1}$.
By using the fact that $\frac{n}{\log n}$ is an increasing function of n, and that the relation $\frac{y}{\log y} \ge x$ is implied by $y \ge x \cdot \log x \cdot \log\log x$, we have the following sufficient condition on n:

When $\alpha > 1$ we assume that

$$n \;\ge\; (4\alpha + 2)^{2\alpha+2} \cdot (\log(4\alpha + 2))^2$$

We use the threshold

$$\lambda_{n,\delta} = C_\varphi\, b\, \left(1 + 2\sqrt{(1+\delta)\ln 2}\right) \sqrt{\frac{\log n}{n}}$$

The first step of the wavelet shrinkage algorithm is DWT, which maps $(y_1,\ldots,y_n)$ to $\sqrt{n}\,(c^{(y)}_{J_0,0},\ldots,c^{(y)}_{J_0,2^{J_0}-1},d^{(y)}_{J_0,0},\ldots,d^{(y)}_{J_0,2^{J_0}-1},\ldots,\ldots,\ldots,d^{(y)}_{J-1,0},\ldots,d^{(y)}_{J-1,2^{J-1}-1})$, where $n=2^J$. Since $y_i=f_i+e_i$ and the DWT is linear we have

$$c^{(y)}_{J_0,k}=c^{(f)}_{J_0,k}+c^{(e)}_{J_0,k},\qquad 0\le k<2^{J_0},$$

and

$$d^{(y)}_{j,k}=d^{(f)}_{j,k}+d^{(e)}_{j,k},\qquad J_0\le j<J,\quad 0\le k<2^j,$$

where $c^{(f)}_{J_0,k}$, $d^{(f)}_{j,k}$ and $c^{(e)}_{J_0,k}$, $d^{(e)}_{j,k}$ are the wavelet coefficients for $(f_1,\ldots,f_n)$ and $(e_1,\ldots,e_n)$, respectively.

The second step of wavelet shrinkage is thresholding. We shall prove our result for soft thresholding. But in our proofs it will be easy to see that our results will hold for hard thresholding too. For soft thresholding, we have

$$\tilde{d}_{j,k}=\begin{cases} d^{(y)}_{j,k}-\lambda_{n,\delta} & \text{if } d^{(y)}_{j,k}>\lambda_{n,\delta}\\ 0 & \text{if } |d^{(y)}_{j,k}|\le\lambda_{n,\delta}\\ d^{(y)}_{j,k}+\lambda_{n,\delta} & \text{if } d^{(y)}_{j,k}<-\lambda_{n,\delta}\end{cases}$$

The last step of wavelet shrinkage is the inverse of DWT which yields $\tilde{y}=(\tilde{y}_1,\ldots,\tilde{y}_n)$. If we let

$$\tilde{y}(x)=\sum_{k=0}^{2^{J_0}-1}c^{(y)}_{J_0,k}\,\varphi_{J_0,k}(x)+\sum_{j=J_0}^{J-1}\sum_{k=0}^{2^{j-1}}\tilde{d}_{j,k}\,\psi_{j,k}(x),\tag{2}$$

then we obtain $\tilde{y}_i=\tilde{y}(\frac{i}{n})$ for $i=1,\ldots,n$.

3.2 Application of Talagrand's theorem

Let W be the orthogonal matrix that represents the DWT. Let $A\subseteq\Omega^n$ be as above. For any $\delta>0$ we define the following subset of Ω^n:

$$B_\delta=\left\{\omega'\in\Omega^n:\ (\forall\ell\in[1,n]),\ \inf_{\omega\in A}\left|\sum_{i=1}^{n}W_{\ell,i}(e_i(\omega')-e_i(\omega))\right|<\ 2b\sqrt{(1+\delta)\ln n}\ \right\}.$$

Lemma 3.4 *For all $\omega'\in B_\delta$ and all $k\in[0,2^{J_0}-1]$:* $|c^{(e(\omega'))}_{J_0,k}|\le\lambda_{n,\delta}$.

For all $j\in[J_0,J-1]$ and $k\in[0,2^j-1]$: $|d^{(e(\omega'))}_{j,k}|\le\lambda_{n,\delta}$.

Proof: By the definition of B_δ, for every $\omega' \in B_\delta$ there exists $\omega \in A$ such that

$$\sqrt{n}\ |c_{J_0,k}^{(e(\omega))} - c_{J_0,k}^{(e(\omega'))}| \le b2\sqrt{(1+\delta)\ln n}$$

and

$$\sqrt{n}\ |d_{j,k}^{(e(\omega))} - d_{j,k}^{(e(\omega'))}| \le b2\sqrt{(1+\delta)\ln n}$$

The Lemma then follows from Lemma 3.3. $\square$

For the following theorem we use the threshold $\lambda_{n,\delta}$ as above; we let $n_0 = 2^9$ when $0 < \alpha \le 1$, and $n_0 = (4\alpha+2)^{2\alpha+2} \cdot (\log(4\alpha+2))^2$ when $\alpha > 1$.

Lemma 3.5 *When* $n \ge n_0$, $\quad P(B_\delta) > 1 - \frac{9}{n^{1+\delta}}$.

Proof: We first prove that

$$\{\omega' \in \Omega^n \ :\ d_T(\omega', A) < 2\sqrt{(1+\delta)\ln n}\} \ \subseteq\ B_\delta.$$

Recall the definition
$d_T(\omega', A) =$
$\sup\{\ \inf\{\sum_{i=1}^n \beta_i \cdot I(\omega_i' \ne \omega_i) : (\omega_1, \dots, \omega_n) \in A\} : (\beta_1, \dots, \beta_n) \in \mathbb{R}^n,\ \sum_{i=1}^n \beta_i^2 = 1\}$.
We will choose the following n vectors for $\beta = (\beta_1, \dots, \beta_n)$ in the above formula:
$(|W_{1,\ell}|, \dots, |W_{n,\ell}|)$, for $\ell = 1, \dots, n$.
Since W is orthogonal all its row vectors have unit length. For all $\omega' \in \Omega^n$, $\omega = (\omega_1, \dots, \omega_n) \in A$, and $1 \le \ell \le n$, we have:
$|\sum_{i=1}^n W_{i,\ell}(e_i(\omega') - e_i(\omega))|$
$\le\ b\sum_{i=1}^n |W_{i,\ell}| \cdot I(e_i(\omega') \ne e_i(\omega))$
$\le\ b\sum_{i=1}^n |W_{i,\ell}| \cdot I(\omega' \ne \omega)$.
(The last inequality follows from the fact that $I(e_i(\omega') \ne e_i(\omega)) \le I(\omega' \ne \omega)$, because $e_i(\omega') \ne e_i(\omega)$ implies $\omega' \ne \omega$.)
Hence, for all $\omega' \in \Omega^n$ and $1 \le \ell \le n$,
$\inf\{|\sum_{i=1}^n W_{i,\ell}(e_i(\omega') - e_i(\omega))| : \omega \in A\}$
$\le\ \inf\{\sum_{i=1}^n |W_{i,\ell}| \cdot I(\omega' \ne \omega)\, b : \omega \in A\}$
$=\ b\ \inf\{\sum_{i=1}^n |W_{i,\ell}| \cdot I(\omega' \ne \omega) : \omega \in A\}$.
Therefore, if $d_T(\omega', A) \le 2\sqrt{(1+\delta)\ln n}$ then for all $1 \le \ell \le n$,
$\inf\{|\sum_{i=1}^n W_{i,\ell}(e_i(\omega') - e_i(\omega))| : \omega \in A\}\ \le\ b\,2\sqrt{(1+\delta)\ln n}$.
This means that $\omega' \in B_\delta$, and this proves that
$\{\omega' \in \Omega^n\ :\ d_T(\omega', A) < 2\sqrt{(1+\delta)\ln n}\}\ \subseteq\ B_\delta$.
Hence, $P(B_\delta)\ \ge\ P(\{\omega' \in \Omega^n : d_T(\omega', A) < 2\sqrt{(1+\delta)\ln n}\})$.
By Talagrand's theorem this is $\ge\ 1 - \exp(-(1+\delta)\ln 2) \cdot \frac{1}{P(A)}\ >\ 1 - \frac{9}{n^{1+\delta}}$. $\square$

Lemma 3.6 *For all* $\omega' \in B_\delta$ *we have:*

(1) *When* $J_1 \le j < J$, $0 \le k < 2^j$, $\quad |\tilde{d}_{j,k}(\omega') - d^{(f)}_{j,k}| \;\le\; |d^{(f)}_{j,k}| \;\le\; C_\varphi M \cdot 2^{-j(\frac{1}{2}+\alpha)}$.

(2) *When* $J_0 \le j < J_1$, $0 \le k < 2^j$, $\quad |\tilde{d}_{j,k}(\omega') - d^{(f)}_{j,k}| \;\le\; 2\lambda_{n,\delta}$.

Proof: To prove (1), we note first that by (H3), (D3) we have $|d^{(f)}_{j,k}| \le C_\varphi M\, 2^{-j(1/2+\alpha)}$.
To prove the inequality $|d^{(f)}_{j,k} - \tilde{d}_{j,k}| \le |d^{(f)}_{j,k}|$ one considers six cases, according to the possible relative positions of 0, $d^{(f)}_{j,k}$, and $\tilde{d}_{j,k}$. If $0 \le \tilde{d}_{j,k} \le d^{(f)}_{j,k}$, or if $d^{(f)}_{j,k} \le \tilde{d}_{j,k} \le 0$, the inequality is obvious from the order picture. The other four cases are not possible, since they would imply that $|d^{(e(\omega))}_{j,k}| > \lambda_{n,\delta}$, contradicting what we saw a little earlier. This proves (1).

For the proof of (2) we consider two cases. If $\tilde{d}_{j,k} = 0$, $|d^{(y)}_{j,k}| \le \lambda_{n,\delta}$, hence $|d^{(f)}_{j,k} - \tilde{d}_{j,k}| = |d^{(f)}_{j,k}| = |d^{(y)}_{j,k} - d^{(e)}_{j,k}| \le |d^{(y)}_{j,k}| + |d^{(e)}_{j,k}| \le \lambda_{n,\delta} + \lambda_{n,\delta}$. In the second case, $|d^{(y)}_{j,k}| > \lambda_{n,\delta}$, and $|d^{(f)}_{j,k} - \tilde{d}_{j,k}| = |d^{(e)}_{j,k} - \lambda_{n,\delta}| \le \lambda_{n,\delta} + \lambda_{n,\delta}$. This proves the inequality. $\Box$

Theorem 3.7 (Deviation bound for max square error) *For wavelet shrinkage with threshold* $\lambda_{n,\delta}$ *we have for all* $n \ge n_0$*:*

$$P\left(\max_{0\le i\le n}(f_i - \tilde{y}_i)^2 \;\le\; (c_1 + c_2\,\delta)\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}\right) \;\ge\; 1 - \frac{9}{n^{1+\delta}}$$

where c_1 *and* c_2 *depend only on* b, M, *and* α.
As a consequence **(deviation bound for mean square error)**,

$$P\left(\frac{1}{n}\sum_{i=0}^{n}(f_i - \tilde{y}_i)^2 \le (c_1 + c_2\,\delta)\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}\right) \;\ge\; 1 - \frac{9}{n^{1+\delta}}$$

Proof: At the beginning of subsection 2.1 we defined the function $\overline{f}$, and its wavelet coefficients. We have

$$\overline{f}(x) = \sum_{k=0}^{2^{J_0}-1} c^{(f)}_{J_0,k}\varphi_{J_0,k}(x) + \sum_{j=J_0}^{J_1-1}\sum_{k=0}^{2^j-1} d^{(f)}_{j,k}\psi_{j,k}(x) + \sum_{j=J_1}^{J-1}\sum_{k=0}^{2^j-1} d^{(f)}_{j,k}\psi_{j,k}(x),$$

and $f_i = \overline{f}(\frac{i}{n})$ for $1 \le i \le n$.

In connection with the thresholding of y we define the function

$$\tilde{y}(x) = \sum_{k=0}^{2^{J_0}-1} c^{(y)}_{J_0,k}\varphi_{J_0,k}(x) + \sum_{j=J_0}^{J_1-1}\sum_{k=0}^{2^j-1} \tilde{d}_{j,k}\psi_{j,k}(x) + \sum_{j=J_1}^{J-1}\sum_{k=0}^{2^j-1} \tilde{d}_{j,k}\psi_{j,k}(x).$$

By Lemma 3.4 we have for all $\omega' \in B_\delta$:

(0) $\quad |c^{(y)}_{J_0,k} - c^{(f)}_{J_0,k}| = |c^{(e(\omega'))}_{J_0,k}| \;\le\; \lambda_{n,\delta}$

By Lemma 3.6 we have for all $\omega' \in B_\delta$:

(1) $\quad |\tilde{d}_{j,k} - d^{(f)}_{j,k}| \ \le\ |d^{(f)}_{j,k}| \ \le\ C_\varphi M \cdot 2^{-j(\frac{1}{2}+\alpha)} \qquad$ for $J_1 \le j < J,\ 0 \le k < 2^j$

(2) $\quad |\tilde{d}_{j,k} - d^{(f)}_{j,k}| \ \le\ 2\lambda_{n,\delta} \qquad$ for $J_0 \le j < J_1,\ \ 0 \le k < 2^j$.

Let us first deal with the case of Haar wavelets (when $\alpha \le 1$). For a given j, the supports of different Haar wavelets do not overlap. Therefore, for all $x \in\,]0,1]$ there exist K_1 and $K(j)$ such that

$|\tilde{f}(x) - \tilde{y}(x)| \ \le$

$$|c^{(y)}_{J_0,K_1} - c^{(f)}_{J_0,K_1}| \cdot 2^{J_0/2} + \ \sum_{j=J_0}^{J_1-1} |\tilde{d}_{j,K(j)} - d^{(f)}_{j,K(j)}| \cdot 2^{j/2} + \ \sum_{j=J_1}^{J-1} |\tilde{d}_{j,K(j)} - d^{(f)}_{j,K(j)}| \cdot 2^{j/2}$$

This and (0), (1), (2) imply for all $x \in\,]0,1]$:

$$|\tilde{f}(x) - \tilde{y}(x)| \ \le\ C_1 \cdot \left(\tfrac{\log n}{n}\right)^{\frac{\alpha}{1+2\alpha}} + C_2 \cdot \left(\tfrac{\log n}{n}\right)^{\frac{\alpha}{1+2\alpha}} + C_3 \cdot \left(\tfrac{\log n}{n}\right)^{\frac{\alpha}{1+2\alpha}}$$
$$= (c_1' + c_2'\sqrt{1+\delta}) \cdot \left(\tfrac{\log n}{n}\right)^{\frac{\alpha}{1+2\alpha}}$$

Letting $x = \frac{i}{n}$ $(1 \le i \le n)$ we obtain for all $\omega' \in B_\delta$:

$$|f_i - \tilde{y}_i(\omega')| = |\tilde{f}(\tfrac{i}{n}) - \tilde{y}(\tfrac{i}{n})| \ \le\ (c_1' + c_2'\sqrt{1+\delta}) \cdot \left(\tfrac{\log n}{n}\right)^{\frac{\alpha}{1+2\alpha}}$$

In the Haar case the theorem follows from this and the fact that $P(B_\delta) > 1 - \frac{9}{n^{1+\delta}}$ (when $n \ge n_0$).

For wavelets on the interval (when $\alpha > 1$, and the number of vanishing moments is $N = \lceil \alpha \rceil$), there are never more than $2N$ wavelets that overlap (for a given j). Indeed, in the above sums we have for each j and each x: $\ 0 \le 2^j x - k \le 2N - 1$. (Other values of k would place the argument $2^j x - k$ of the wavelet functions outside of the support and would ence only produce zero-terms in the sums.) Hence k only needs to range from $\lceil 2^j x \rceil - 2N + 1$ through $\lceil 2^j x \rceil$, which corresponds to $2N$ values of k.

Hence, the same calculation as for Haar wavelets applies, except that the constants C_1, C_2, C_3, c_1', c_2' need to be multiplied by $2N$. $\square$

Appendix

Proof of Lemma 2.1

Properties (H1) and (H2) follow from a direct calculation based on the exact formulas for the Haar wavelets $\varphi_{j,k}$ and $\psi_{j,k}$.

$$c^{(e)}_{j,k} = \int_0^1 \overline{e}(x)\varphi_{j,k}(x)dx = 2^{j/2} \int_{k2^{-j}}^{(k+1)2^{-j}} \overline{e}(x)dx =$$
$$\sum_{i=k2^{J-j}}^{(k+1)2^{J-j}-1} e_{i+1} 2^{-J} = 2^{-J+j/2} \sum_{i=0}^{2^{J-j}-1} e_{i+1+k2^{J-j}}.$$

The calculation for (H2) is similar. The same calculation as for (H2) will give for $\overline{f}$:

$$d^{(f)}_{j,k} = 2^{-J-1+j/2} \sum_{i=0}^{2^{J-j}-1} (f(i+1+k2^{J-j}) - f(i+1+(k+\tfrac{1}{2})2^{J-j})).$$

Then we use the Hölder condition $\ |f(i+1+k2^{J-j}) - f(i+1+(k+\frac{1}{2})2^{J-j})| \ \le\ M\,(\frac{1}{2}2^{J-j})^\alpha$. $\square$

Proof of Lemma 2.2

Property (D1) follows from a direct calculation:

$$c_{j,k}^{(e)} = \int_0^1 \overline{e}(x)\varphi_{j,k}(x)dx = \sum_{i=0}^{n-1} e_i \int_{\frac{i}{n}}^{\frac{i+1}{n}} \varphi_{jk}(x)dx$$

where we denote the functions $\varphi_{jk}^{\text{left}}$ by $\varphi_{j,2^j-2N+k}$, and $\varphi_{jk}^{\text{right}}$ by $\varphi_{j,2^j-N+k}$.
For the φ_{jk} "in the middle" of the interval we have

$$\int_{\frac{i}{n}}^{\frac{i+1}{n}} \varphi_{jk}(x)dx = 2^{j/2}\int_{i2^{-J+j}-k}^{(i+1)2^{-J+j}-k} \varphi(t)\, 2^{-j}dt = 2^{j/2}2^{-J}\alpha_{ijk}$$

by the the mean-value theorem, for some numbers α_{ijk} with $|\alpha_{ijk}| \leq \sup_{[0,1]}|\varphi|$.
For the $\varphi_{j,2^j-2N+k}$ "at the left end" of the interval,

$$\int_{\frac{i}{n}}^{\frac{i+1}{n}} \varphi_{jk}^{\text{left}}(x)dx = \int_{\frac{i}{n}}^{\frac{i+1}{n}} \sum_{s=0}^{2N-1}(-s)^k\,\varphi(2^jx+s)dx = \sum_{s=0}^{2N-1}(-s)^k \int_{i2^{-J+j}+s}^{(i+1)2^{-J+j}+s} \varphi(y)\, 2^{-j}dy$$
$$= \sum_{s=0}^{2N-1}(-s)^k 2^{-j}2^{-J+j}\gamma_{ijs}$$

by the mean-value theorem, for some numbers γ_{ijs} with $|\gamma_{ijs}| \leq \sup_{[0,1]}|\varphi|$. By taking

$$\alpha_{ijk} = 2^{-j/2}\sum_{s=0}^{2N-1}(-s)^k\gamma_{ijs}$$

we obtain (D1). At the left end, $k \leq N$, so $|\alpha_{ijk}| \leq 2N(2N-1)^N \cdot \sup|\varphi|$.

The scaling functions "at the right end" of the interval are handled in a similar way. The calculation for (D2) is similar. (D3) follows from the wavelet characterization of Hölder classes ([5], page 299, and [10]). $\square$

References

[1] R. Averkamp, Ch. Houdré, "Wavelet Thresholding for Non (Necessarily) Gaussian Noise: Idealism", preprint (\protect\vrule width0pt\protect\href{http://www.math.gatech.edu\string~houdre/}{htt

[2] R. Averkamp, Ch. Houdré, "Wavelet Thresholding for Non (Necessarily) Gaussian Noise: Functionality", preprint (\protect\vrule width0pt\protect\href{http://www.math.gatech.edu\string~houdre/}{htt

[3] L.D. Brown, M.G. Low, "Superefficiency and lack of adaptability in functional estimation", manuscript.

[4] A. Cohen, I. Daubechies, B. Jawerth, P. Vial, "Multiresolution analysis, wavelets and fast algorithms on an interval", *Comptes Rendus de l'Académie des Science de Paris*, t. 316, Série I (1993) 417-421.

[5] I. Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics (1992).

[6] D. Donoho, I. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika* 81(3) (1994) 425-455.

[7] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, "Wavelet shrinkage: Asymptopia?", *Journal of the Royal Statistics Society* series B, 57(2) (1995) 301-369.

[8] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *Journal of the American Statistical Association* 58 (1965) 13-30.

[9] O.V. Lepskii, "On one problem of adaptive estimation on white Gaussian noise", *Teor. Veoryatnost. i Premenen.* 35 (1990) 459-470 [Russian]. *Theory of Probability and Applications* 35 (1990) 454-466 [English].

[10] Y. Meyer, *Wavelets and Operators*, Cambridge University Press (1992).

[11] J.M. Steele, *Probability Theory and Combinatorial Optimization*, Society for Industrial and Applied Mathematics (1997).

[12] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces", *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques* 81 (1995) 73-205.

[13] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley (1999).